

# MIAGE\_qcm1\_2023

1. [ACP] Sur le cercle des corrélations,

Si deux flèches représentant respectivement les variables A et B sont très proches du bord du cercle, et pointant dans des directions opposées, alors

- (A) B a tendance à diminuer quand A augmente
- (B) B a tendance à augmenter quand A augmente
- (C) On ne peut rien dire
- (D) Il y a une grande proportion d'individus avec à la fois une petite valeur de A et une petite valeur de B

2. [ACP] Combien d'axes retenir au minimum pour conserver 80% de l'information ?

> res.pca\$eig

	eigenvalue	percentage of variance	cumulative percentage of variance
comp 1	2.8445114	35.556392	35.55639
comp 2	1.3053388	16.316735	51.87313
comp 3	1.2197992	15.247489	67.12062
comp 4	0.9365384	11.706730	78.82735
comp 5	0.6970332	8.712915	87.54026
comp 6	0.4722925	5.903656	93.44392
comp 7	0.3086008	3.857510	97.30143
comp 8	0.2158858	2.698573	100.00000

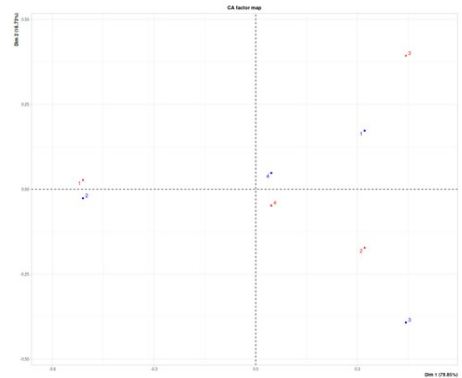
- (A) 5
- (B) 3
- (C) 4
- (D) 6

3. [AFC] Si deux profils lignes sont superposés sur la représentation graphique (simultanée), alors

- (A) Ces profils sont proportionnels
- (B) Ces profils sont égaux
- (C) On ne peut rien dire, en revanche pour des colonnes il y aurait proportionnalité
- (D) On ne peut rien dire, en revanche pour des colonnes il y aurait égalité

4. [AFC] Sur la représentation graphique simultanée ci-dessous,

- A) 3 rouge est du côté de 1 bleu, donc ces deux modalités ont tendance à s'associer
- B) 1 rouge est très proche de 2 bleu : ces modalités sont en association
- C) 4 rouge est très proche de 4 bleu, de plus ces deux modalités sont représentées proches du centre. Elles sont donc en association
- D) 4 rouge est très proche du centre, c'est donc une modalité rare



5. [ACP] Cochez l'affirmation juste...

- A) Si sur le cercle des corrélations deux flèches sont très courtes (très loin du bord), alors on ne peut rien dire concernant l'angle formé par les variables correspondantes (non projetées)
- B) Si sur le cercle des corrélations deux flèches sont très courtes (très loin du bord), alors l'angle formé par les variables correspondantes (non projetées) est très petit
- C) Le nuage des individus indique les corrélations entre variables
- D) Le nuage des individus s'affiche toujours en projection sur les deux premiers axes

6. [Clustering] Concernant l'algorithme des k-means...

- A) On met à jour les centres en effectuant la moyenne des individus de chaque groupe, composante par composante
- B) On met à jour les centres en effectuant la somme des individus de chaque groupe, composante par composante
- C) Le résultat est le même quelle que soient les centres initiaux
- D) Les centres appartiennent forcément au jeu de données

7. [Clustering] Concernant l'algorithme des k-medoids...

- A) On met à jour les centres en minimisant la somme des distances d'un point à tous les autres du même groupe
- B) On met à jour les centres en minimisant la somme des carrés des distances d'un point à tous les autres du même groupe
- C) Le résultat est le même quelle que soient les centres initiaux
- D) Les centres n'appartiennent pas forcément au jeu de données

**8.** [Clustering] Concernant l'algorithme CAH...

- A** Le dendrogramme résume graphiquement la hiérarchie des partitions obtenues au fil de l'algorithme, avec les coûts de fusion
- B** Le dendrogramme ne donne que la partition finale avec une valeur de  $k$  fixée
- C** Les seules distances inter-groupes utilisables par l'algorithme sont single-linkage et complete-linkage
- D** Le choix de la distance inter-groupes n'a aucune incidence sur la hiérarchie des partitions, elle n'influe que sur les coûts de fusion

**9.** [Clustering] Concernant le graphe des  $k$  plus proches voisins ( $k$ -PPV)...

- A** Le graphe des  $k$ -PPV mutuels est obtenu à partir du graphe des  $k$ -PPV orienté, en supprimant les arêtes non bi-directionnelles
- B** La distance du plus court chemin calculée dans le graphe des  $k$ -PPV est toujours inférieure à la distance euclidienne
- C** On choisit toujours  $k = \sqrt{n}$ , c'est la valeur optimale
- D** On choisit toujours  $k = n^{1/3}$ , c'est la valeur optimale

**10.** [Clustering] Le clustering spectral est un algorithme approximant un problème discret défini dans un graphe. Ce problème donne une fonction objectif raisonnable pour la recherche de groupes

- T** True
- F** False